

# Aproximaciones a la imputación de ingresos desde los estudios de análisis de clase. Una propuesta a partir del uso de técnicas de análisis factorial.

María Clara Fernández Melián y José Javier Rodríguez de la Fuente.

Cita:

María Clara Fernández Melián y José Javier Rodríguez de la Fuente (2019). *Aproximaciones a la imputación de ingresos desde los estudios de análisis de clase. Una propuesta a partir del uso de técnicas de análisis factorial.*

Dirección estable: <https://www.aacademica.org/joserodriguez/51>

ARK: <https://n2t.net/ark:/13683/pq7B/drv>



Esta obra está bajo una licencia de Creative Commons.  
Para ver una copia de esta licencia, visite  
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

*Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.*



# INCASI Working Paper Series

2019, No. 8

 **INCASI** *International Network for  
Comparative Analysis of Social Inequalities*



**Aproximaciones a la imputación de ingresos  
desde los estudios de análisis de clase.  
Una propuesta a partir del uso de técnicas de  
análisis factorial**

**María Clara Fernández Melián  
José Javier Rodríguez de la Fuente**



European  
Commission

Horizon 2020  
European Union funding  
for Research & Innovation

Marie Skłodowska-Curie Actions (MSCA)  
Research and Innovation Staff Exchange (RISE)  
H2020-MSCA-RISE-2015  
GA-691004





**UAB**

Universitat Autònoma de Barcelona

**iet** Institut  
d'Estudis  
del Treball**gret** Grup de Recerca  
en Educació i  
Treball**INCASI** *International Network for  
Comparative Analysis of Social Inequalities*

# Aproximaciones a la imputación de ingresos desde los estudios de análisis de clase. Una propuesta a partir del uso de técnicas de análisis factorial.

María Clara Fernández Melián<sup>1</sup>  
José Javier Rodríguez de la Fuente<sup>2</sup>

<sup>1</sup> Licenciada en Sociología. Becaria doctoral del Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET). Instituto de Investigaciones Gino Germani (IIGG), Universidad de Buenos Aires (UBA) / Universidad Nacional de Tres de Febrero (UNTREF) Argentina  
[email@email.com](mailto:email@email.com)

<sup>2</sup> Licenciado en Sociología. Becario doctoral del Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET). Instituto de Investigaciones Gino Germani (IIGG), Universidad de Buenos Aires (UBA). Argentina  
[email@email.com](mailto:email@email.com)

**UBA**  
Universidad de Buenos Aires  
*Argentina virtus robor et studium***European  
Commission**Horizon 2020  
European Union funding  
for Research & Innovation

Marie Skłodowska-Curie Actions (MSCA)  
Research and Innovation Staff Exchange (RISE)  
H2020-MSCA-RISE-2015  
GA-691004

INCASI Working Paper Series is an online publication under *Creative Commons* license. Any person is free to copy, distribute or publicly communicate the work, according to the following conditions:



*Attribution.* All CC licenses require that others who use your work in any way must give you credit the way you request, but not in a way that suggests you endorse them or their use. If they want to use your work without giving you credit or for endorsement purposes, they must get your permission first.



*NonCommercial.* You let others copy, distribute, display, perform, and (unless you have chosen NoDerivatives) modify and use your work for any purpose other than commercially unless they get your permission first.



*NoDerivatives.* You let others copy, distribute, display and perform only original copies of your work. If they want to modify your work, they must get your permission first.

There are no additional restrictions. You cannot apply legal terms or technological measures that legally restrict doing what the license allows.

This working paper was elaborated in the context of INCASI Network, a European project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie GA, No. 691004, and coordinated by Dr. Pedro López-Roldán. This article reflects only the author's view and the Agency is not responsible for any use that may be made of the information it contains.

Digital edition: <https://ddd.uab.cat/record/204785>

Dipòsit Digital de Documents  
Bellaterra, Cerdantola del Vallès (Barcelona)  
Universitat Autònoma de Barcelona



# Aproximaciones a la imputación de ingresos desde los estudios de análisis de clase. Una propuesta a partir del uso de técnicas de análisis factorial<sup>1</sup>

María Clara Fernández Melián  
José Javier Rodríguez de la Fuente

## *Abstract*

El propósito del trabajo es analizar las implicancias de la utilización de distintos métodos de imputación de datos en relación a la variable sobre ingresos. En particular, interesa describir y aplicar diferentes métodos y dar cuenta de las similitudes y diferencias entre ellos. De esta manera, se espera contribuir a la problemática de la falta de respuesta por parte de los informantes a las preguntas sobre ingresos de las encuestas. Específicamente, se parte del análisis de la estructura de clases, por lo cual, se consideran los siguientes métodos de imputación: por un lado, la asignación de la mediana de ingresos de la clase social a la que pertenece el encuestado y por el otro, la técnica multivariada de análisis de correspondencias múltiples. Ésta última permite observar la distribución y ubicación en el “espacio social multivariado” de aquellas unidades de análisis sin información sobre ingresos en relación a diversas dimensiones tales como la clase social, el nivel educativo, el consumo, la zona de residencia y también en relación al ingreso declarado por los restantes informantes. Se utilizará como fuente de información la encuesta “Movilidad social y opiniones sobre la sociedad actual” realizada entre 2012 y 2013 en la Ciudad Autónoma de Buenos Aires, llevada a cabo por el grupo de estudios sobre “Desigualdad y movilidad social”, del Instituto de Investigaciones Gino Germani, Universidad de Buenos Aires.

## *Keywords*

Análisis de clase, ingresos, imputación, análisis factorial de correspondencias múltiples

## *Index*

1. Introducción. 2. Metodología. 3. Caracterización de casos con información de ingreso faltante. 4. Imputación de ingreso a partir de la clase social. 5. Imputación a partir de la técnica de análisis de correspondencias múltiples. 6. Consideraciones finales. 7. Bibliografía. 8. Anexo.

## 1. Introducción

Este trabajo constituye una primera aproximación a la cuestión de la imputación de valores faltantes en una base de datos y realiza una propuesta para el tratamiento de los mismos. A partir de un documento previo en el que se

pretendía caracterizar la relación existente entre el posicionamiento de clase y el nivel de riqueza acumulado en el hogar (mediante un índice que contemplaba el ingreso, el consumo y la vivienda)<sup>2</sup> se hizo necesario resolver el problema de la falta de respuesta en relación a los ingresos. En particular, se tenían 174 casos con

<sup>1</sup> Los autores agradecen los comentarios recibidos a este documento presentado en la sesión del RC55 “*Imputation and Social Indicators: The Use of Factor Analysis for Imputing Missing Data*” del III ISA Forum, realizado en Viena, Austria el 14 de julio de 2016.

<sup>2</sup> Para más información ver Clemenceau, Fernández Melián, y Rodríguez de la Fuente (2016).



información faltante en dicha variable (aproximadamente un 25% de los casos).

Según Cochran (1977) el problema de falta de respuesta en las encuestas de hogares se presenta por:

- a) Cobertura. No se pudo ubicar algunas de las unidades seleccionadas debido a problemas de acceso.
- b) Localización. Cuando no fue posible localizar a ningún miembro del hogar durante la visita a la vivienda seleccionada.
- c) Informante inadecuado. Cuando la persona entrevistada no está en posibilidad de proporcionar la información que se le demanda.
- d) Rechazo. Cuando los hogares seleccionados se niegan a participar en la encuesta (Cochran, 1977, en Galván y Medina, 2007: 9).

La imputación de datos es un “proceso de estimación de valores ausentes basado en valores válidos de otras variables o casos de la muestra” (Pérez López, 2004: 46) con el objetivo de no reducir el tamaño de la muestra lo que sucedería en caso de excluir dichos casos del análisis (procedimiento conocido como *listwise*). Si bien, como indican Galván y Medina (2007: 21), el *listwise* es el procedimiento utilizado más frecuentemente, el mismo asume que la falta de datos sigue un patrón totalmente aleatorio, introduciendo posteriores sesgos al eliminar casos que posiblemente provengan de un muestreo probabilístico, en el cual las unidades fueron elegidas, en algunas de las etapas, en forma aleatoria. Algunas de las formas de imputación son la sustitución de datos ausentes por la media o la mediana, el método de sustitución por un valor constante, el método de imputación por interpolación lineal, el método de imputación por regresión y el método de imputación múltiple (2004). Otra variante es formar categorías a partir de covariables correlacionadas con la variable de interés e imputar los datos faltantes con observaciones de la submuestra con la cual comparte características (Acock y Demo, 1994). Hace unos

años, el muestreo de tipo *Hot-deck* (en todas sus variantes) se ha utilizado para la imputación de datos faltantes en censos y encuestas de hogares, debido a su carácter no paramétrico y su capacidad para generar el reemplazo a partir de una selección aleatoria de valores observados (donantes), no introduciendo sesgos en la varianza del estimador (Galván y Medina, 2007: 27).

El objetivo de este trabajo es analizar dos métodos de imputación de ingresos a partir de la consideración de que el ingreso de un individuo puede asimilarse al que poseen quienes presentan características sociales y ocupacionales similares. En este sentido, trabajando con la variable de ingresos medida en intervalos, en primer lugar, se realiza un ejercicio de imputación, en el que se le otorga la mediana de ingresos en base a la clase social de pertenencia bajo el supuesto que cada individuo tiene condiciones económicas similares a los restantes integrantes de cada grupo. En segundo lugar, como contraste, se aplicará la técnica de análisis de correspondencias múltiples donde incluiremos distintas variables relevantes para formar grupos diferenciados a partir de los cuales realizar la imputación de los ingresos.

## 2. Metodología

Se utiliza la “Encuesta sobre movilidad social y opiniones sobre la sociedad actual” del año 2012-2013 realizada a través del proyecto FONCyT “Tendencias y transformaciones en la estructura social: El impacto de los procesos de movilidad social en los horizontes de consumo y la participación política. Un análisis de la Región Metropolitana de Buenos Aires. 2003 – 2011”<sup>3</sup>.

La muestra es de tipo probabilística estratificada proporcional y consta de 700 casos. Los estratos se definieron a partir de la composición barrial según Necesidades Básicas Insatisfechas (NBI), los casos fueron seleccionados sistemáticamente

<sup>3</sup> La misma fue coordinada por el Dr. Eduardo Chávez Molina del Instituto de Investigaciones Gino Germani de la Universidad de Buenos Aires y realizada en el marco de un proyecto FONCyT (Fondo para la Investigación Científica y Tecnológica) de la Agencia

Nacional de Promoción Científica y Tecnológica dependiente del Ministerio de Ciencia, Tecnología e Innovación Productiva.

y la asignación fue proporcional por sexo, edad y barrio.

El universo de análisis estuvo conformado por personas mayores a 30 años, ocupadas, que residían en la Ciudad Autónoma de Buenos Aires. Lo que se busca con este rango es captar a individuos que, con mayor probabilidad, se encuentren en una etapa de madurez ocupacional (Echeverría Zabalza, 1999; Erikson y Goldthorpe, 1992), es decir, una edad en la que normalmente los sujetos ya han recorrido la mayor parte de su carrera ocupacional, o al menos están ingresando al período de consolidación laboral.

El abordaje de la estructura de clases parte de la utilización del esquema de clasificación de Susana Torrado (1994, 1998) debido a que el mismo está elaborado para el estudio de la sociedad argentina en particular y ya ha sido empleado en sucesivas investigaciones (Pla, 2016; Sacco, 2016; Pla, Rodríguez y Fernández, 2016). Torrado (1998) construye un nomenclador de Condición Socio Ocupacional (CSO) para definir estratos ocupacionales. Como el objetivo principal de la construcción del nomenclador fue estudiar la estructura de clases, un punto central fue poder obtener e identificar estratos ocupacionales que revistan cierta homogeneidad interna. Además de ser agregados estadísticos con gran frecuencia, estos estratos debían funcionar también como grupos sociales “con cierta identidad” diferenciados entre sí en función de los modos en los cuales los agentes se insertan en los procesos productivos y también, en consecuencia, con condiciones de vida disímiles. Es decir, que a medida que se ascendiese en la jerarquía de la estructura social, las condiciones de vida de los agentes deberían ser mejores, y viceversa, a medida que se descendiese socialmente, las condiciones tenderían a empeorar. Para la construcción de dicho esquema se utilizan las variables de categoría de ocupación (empleador, cuenta propia, empleado, trabajador sin salario o ayuda familiar y servicio doméstico), tamaño del establecimiento (con un umbral de 5 personas que separa a los establecimientos grandes de los pequeños) y ocupación (Clasificador Internacional Uniforme de Ocupaciones – CIUO 08).

A continuación, presentamos el esquema de clases sociales (Cuadro N° 1), propuesto por Torrado (1994; 1998), en dos modalidades: en primer lugar, en su formato ampliado y desagregado de doce clases y en segundo lugar, en un formato de siete clases. En este caso se optó por utilizar el esquema de siete clases.

**Tabla 1. Versión completa y versión colapsada del esquema de clases propuesto por Torrado.**

Versión completa	Versión colapsada
	Siete clases
I. Directores de empresas	Directores de empresas (I)
II. Profesionales en función específica autónomos	Profesionales en función específica (II y III)
III. Profesionales en función específica asalariados	
IV. Propietarios de pequeñas empresas	Propietarios de pequeñas empresas y pequeños productores autónomos (IV y V)
V. Pequeños productores autónomos	
VI. Cuadros técnicos y asimilados	Cuadros técnicos y asimilados (VI)
VII. Empleados administrativos y vendedores	Empleados administrativos y vendedores (VII)
VIII. Trabajadores especializados autónomos	Trabajadores calificados (VIII y IX)
IX. Obreros calificados	
X. Obreros no calificados	Trabajadores no calificados (X, XI y XII)
XI. Peones autónomos	
XII. Empleados domésticos	
Sin especificar la CSO	Sin especificar la CSO

Fuente: elaboración propia en base a esquema de Torrado (1994, 1998), Boado (2008), Sacco (2016) y Pla (2012)

La composición de cada una de las categorías es la siguiente: en primer lugar, los directores o gerentes de empresas incluyen a empleadores o asalariados del sector privado en establecimientos de más de cinco ocupados, que siendo propietarios asumen también tareas de dirección o que siendo asalariados tienen la dirección a su cargo. Los profesionales en función específica comprenden a ingenieros, médicos, economistas, etc. tanto como empleadores, cuenta propia y asalariados de sector privado o público. Los propietarios de pequeñas empresas pertenecen al sector privado



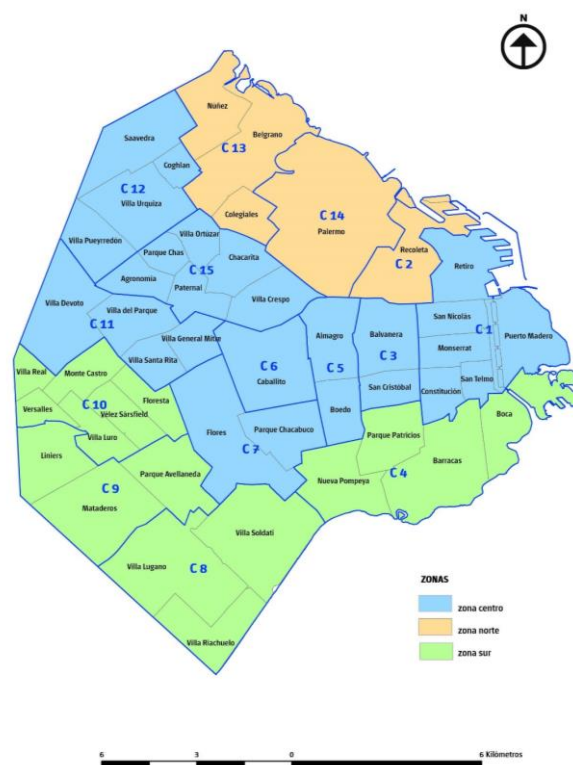
y son empleadores en establecimientos de más de cinco ocupados y que también participan directamente en procesos de trabajo de carácter preponderantemente no manual. Entre los cuadros técnicos y asimilados se cuentan enfermeras, maestros, profesores en la categoría de asalariados de sector privado o público. Los pequeños productores autónomos comprenden a técnicos y afines, comerciantes y vendedores y trabajadores especializados (carpinteros, electricistas, etc.) como empleadores de hasta cinco ocupados o técnicos y comerciantes en la categoría de cuenta propia, todos ellos del sector privado. En el caso de los empleados administrativos y vendedores son asalariados de ambos sectores de actividad como secretarios, cajeros, etc. Los trabajadores especializados autónomos tienen las mismas ocupaciones manuales que los mencionados anteriormente, pero en la categoría cuenta propia. Los obreros son asalariados del sector privado o público, en el caso de los calificados pueden ser carteros, policías, etc. y en los no calificados pueden ser peones, personal de maestranza, etc. Los peones autónomos son trabajadores no especializados en las categorías de empleador o cuenta propia del sector privado. Por último, los empleados domésticos son trabajadores que se desempeñan en hogares particulares.

Respecto a la zona de residencia, se ha utilizado la clasificación elaborada por la Dirección General de Estadísticas y Censos de la CABA, que zonifica a la ciudad de la siguiente forma: Zona Norte: Comunas 2, 13 y 14; Zona Centro: Comunas 1, 3, 5, 6, 7, 11, 12 y 15 y Zona Sur: Comunas 4, 8, 9 y 10 (Mazzeo y Roggi, 2012) (Mapa N° 1).

El consumo se mide a partir de la cantidad de bienes poseídos y comúnmente indagados en relevamientos por encuesta, entre los cuales se considera: TV LCD/LED, notebook, consolas de juego (Playstation, Wii, etc.), heladera con freezer, teléfono celular, conexión a internet, Tv por cable satelital, reproductor Blu-Ray, colchón o somier para cada miembro del hogar, aire acondicionado y cocina con horno. La variable de ingresos fue reconstruida a partir de dos preguntas del cuestionario: una directa relativa al monto de ingreso percibido el mes anterior

(pregunta abierta) y otra indirecta, en el caso de que la primera pregunta no haya sido respondida, para que el entrevistado identifique su nivel de ingresos en una serie de rangos.

**Gráfico 1. Comunas y zonas de la Ciudad Autónoma de Buenos Aires. CABA, 2012**



Fuente: Dirección de Estadísticas y Censos de la Ciudad Autónoma de Buenos Aires.

De esta manera, se pudo recuperar cierta información ya que la pregunta inicial por el monto contó con un gran porcentaje de valores faltantes (tan solo se obtuvo el 24% de nivel de respuesta) a diferencia de la pregunta por rango que mejoró relativamente dicha situación (con el 72% de nivel de respuesta). Las categorías de dicha variable son: 1) hasta \$1.400 (95 USD), 2) \$1.401 a \$3.000 (202 USD), 3) \$3.001 a \$6.000 (405 USD), 4) \$6.001 a \$12.000 (811 USD), 5) \$12.001 a \$24.000 (1.622 USD) y 6) más de \$24.001. Para el presente trabajo, la imputación se realizará con la variable de ingresos construida a partir de rangos debido a las limitaciones mencionadas previamente.

En un primer momento, considerando únicamente los casos con valores válidos en la variable de interés, se imputa la mediana de

ingresos de la clase social de pertenencia a aquellos individuos con información faltante. De este modo, es la posición de clase, la que otorga el criterio imputación a los casos con valores faltantes. Es decir, que si la mediana de ingresos de quienes conforman la clase de directores de empresas se ubica en la categoría de más de \$24.000, ése es el valor asignado en las situaciones con información faltante que pertenecen a dicho grupo. En contraposición, por otro lado, con la utilización del SPAD, se realizó una imputación a través de la utilización del análisis de correspondencias múltiples y el análisis de clasificación (Greenacre, 2007; Le Roux y Rouanet, 2010; López Roldán y Fachelli, 2015). Para este último, se consideran como modalidades activas las variables de clase social, nivel educativo, consumo y zona de residencia y el rango de ingresos como variable suplementaria, es decir, como aquella que no participan activamente en la formación de factores pero que pueden caracterizar o ilustrar los resultados obtenidos (Fachelli, 2012: 78).

De la aplicación de estas técnicas, emergieron agrupamientos que resultaron lo más homogéneos posible en términos de los atributos considerados. Y luego, se asignó la mediana de ingreso del grupo de pertenencia en aquellos casos que no contaran con información sobre ingresos.

### 3. Caracterización de casos

En un primero momento se observaron las características de los casos con información faltante sobre ingresos en relación a la edad, el sexo, el nivel educativo y la categoría de ocupación.

A grandes rasgos, se observa que no hay diferencias significativas entre los casos con información faltante y el resto de la muestra en términos de edad, sexo, nivel educativo ni categoría ocupacional (Tablas 2, 3, 4 y 5).

**Tabla 2. Porcentaje de casos con valor faltante de ingreso y resto según edad**

Edad	Faltantes	Resto
30-39	31,6	37,1
40-49	29,3	27,2
50-59	27,0	25,3
60-69	8,0	8,6
70-79	4,0	1,7
80-89	0,0	0,2
<b>Total</b>	100	100

N = 700. Fuente: elaboración propia en base a Encuesta FONCYT 2012-2013

**Tabla 3. Porcentaje de casos con valor faltante de ingreso y resto según sexo**

Sexo	Faltante	Resto
Varón	50,0	49,0
Mujer	50,0	51,0
<b>Total</b>	100	100

N = 700. Fuente: elaboración propia en base a Encuesta FONCYT 2012-2013

**Tabla 4. Porcentaje de casos con valor faltante de ingreso y resto según nivel educativo**

Nivel educativo	Faltante	Resto
Primaria completa	21,8	25,1
Secundario completo	42,5	40,3
Superior completo o más	35,6	34,6
<b>Total</b>	100	100

N = 700. Fuente: elaboración propia en base a Encuesta FONCYT 2012-2013

**Tabla 5. Porcentaje de casos con valor faltante de ingreso y resto según categoría ocupacional**

Categoría ocupacional	Faltante	Resto
Patrón/Empleador	5,8	8,4
Cuenta Propia	32,9	32,6
Obrero o empleado	55,5	54,5
Trabajador sin salario o ayuda familiar	1,2	0,8
Servicio doméstico	4,6	3,8
<b>Total</b>	100,0	100,0

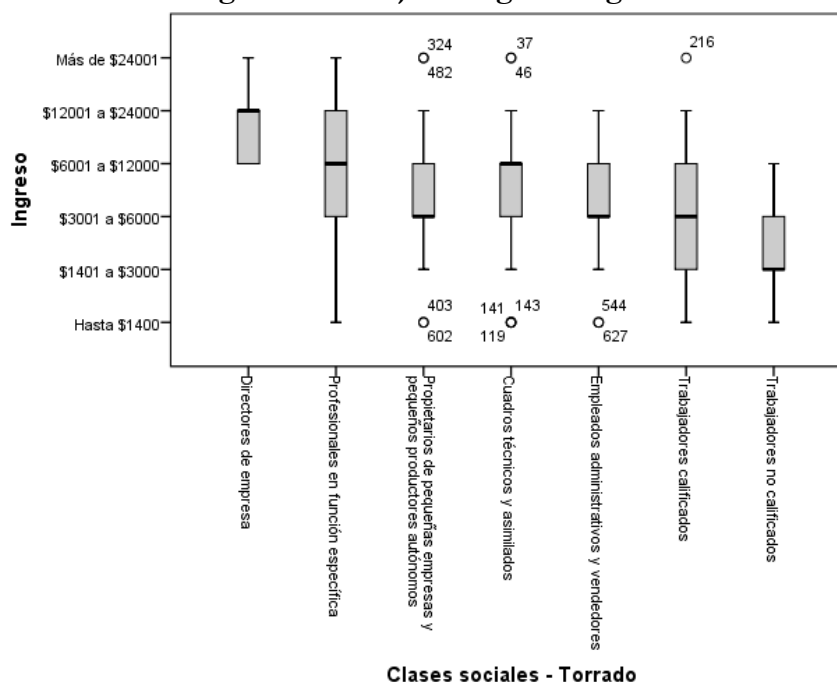
N = 700. Fuente: elaboración propia en base a Encuesta FONCYT 2012-2013

#### 4. Imputación de ingreso a partir de la clase social

En un primer momento se explora la relación entre las variables ingreso y clase social (Gráfico 2) a partir de un diagrama de caja.

En el Gráfico 2 se observa que la distribución de la variable ingreso (en intervalos) tiene valores extremos en cuatro grupos y distribuciones asimétricas, por lo cual se optó por considerar la mediana de ingresos de cada clase social como “valor donante” a aquellos casos en los que es necesario imputar (Tabla 6).

**Gráfico 2. Diagramas de caja de ingreso según clase social**



Nota: hasta \$1.400 (95 USD), \$1.401 a \$3.000 (202 USD), \$3.001 a \$6.000 (405 USD), \$6.001 a \$12.000 (811 USD), \$12.001 a \$24.000 (1.622 USD) y más de \$24.001 (1.622 USD).  
Fuente: elaboración propia en base a Encuesta FONCYT 2012-2013

**Tabla 6. Medianas de ingreso (en intervalos) según clase social**

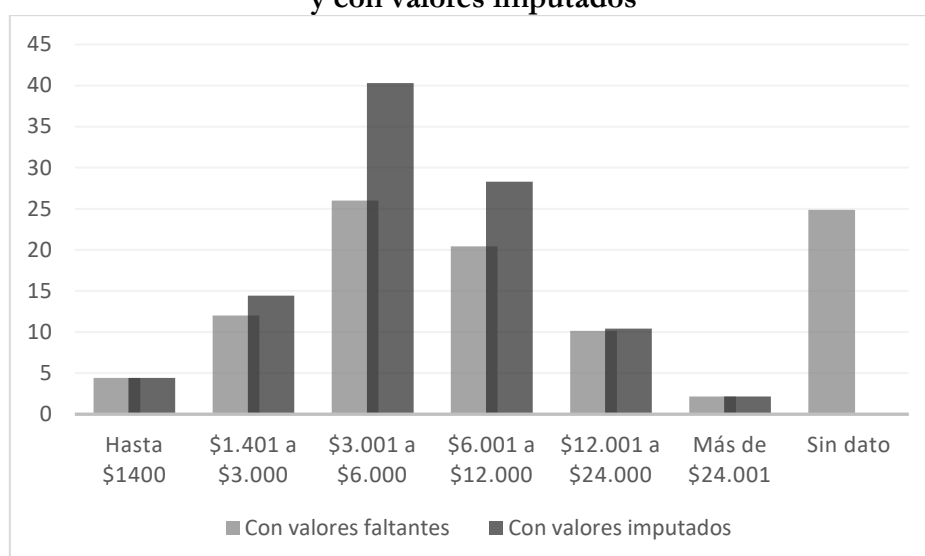
Clase social	Mediana
Directores de empresas	\$12.001 a \$24.000
Profesionales en función específica	\$6.001 a \$12.000
Propietarios de pequeñas empresas y pequeños productores autónomos	\$3.001 a \$6.000
Cuadros técnicos y asimilados	\$6.001 a \$12.000
Empleados administrativos y vendedores	\$3.001 a \$6.000
Trabajadores calificados	\$3.001 a \$6.000
Trabajadores no calificados	\$1.401 a \$3.000
Total	\$3.001 a \$6.000

Fuente: elaboración propia en base a Encuesta FONCYT 2012-2013

Una vez que se realiza la imputación de la mediana de ingresos en base a la clase de pertenencia se observa que los rangos de ingreso que más crecieron a partir de la imputación

fueron los intermedios, de \$3.001 a \$6.000 y de \$6.001 a \$12.000 (Gráfico 3), dando cuenta que el mayor nivel de no respuesta se concentraba en las clases intermedias de la estructura social.

**Gráfico 3. Porcentaje de ingresos según muestra con valores faltantes y con valores imputados**



Nota: hasta \$1.400 (95 USD), \$1.401 a \$3.000 (202 USD), \$3.001 a \$6.000 (405 USD), \$6.001 a \$12.000 (811 USD), \$12.001 a \$24.000 (1.622 USD) y más de \$24.001 (1.622 USD).

Fuente: elaboración propia en base a Encuesta FONCYT 2012-2013

## 5. Imputación a partir de la técnica de análisis de correspondencias múltiples

A continuación, se presentan los resultados obtenidos de realizar el análisis de correspondencias múltiples. Se observa que de las 4 variables y 17 categorías se ha logrado una reducción a un factor que explica más del 70% de la varianza total<sup>4</sup> (Cuadro N° 9 del Anexo). Se suele considerar como criterio la retención del 70% de la varianza en la selección de los factores ya que permite describir los principales elementos de diferenciación de los individuos y aunque se pierda cierta información se gana en significación y parsimonia (López-Roldán, 2012).

A partir de la lectura de las contribuciones absolutas y relativas (Tabla 10 y 11 del Anexo), de los *test values* (Tabla 12 del Anexo) y de los Gráficos 3 y 4, se describen las principales características de los factores que emergen del análisis para luego plantear algunas hipótesis sobre las relaciones que se presentan.

El primer eje factorial explica el 73,6% de la inercia y distribuye a la población en base a diferencias en cuanto a la clase social, la zona de residencia y la percepción de aportes jubilatorios (lo cual puede pensarse como indicador de la formalidad laboral).

Se observa, que en uno de los extremos (puntajes negativos) se encuentran aquellos trabajadores no calificados, sin percepción de jubilación, con nivel educativo hasta primaria completa y que residen en zona sur de la CABA, mientras que en el otro extremo se hallan individuos que son profesionales en función específica y en menor medida directivos, con nivel educativo superior o más, que residen en zona norte y con descuentos jubilatorios. Con respecto a las variables ilustrativas, el eje ubica en un extremo (negativo) a quienes declaran ingresos bajos, menores de \$3.000 y en el otro extremo a quienes perciben ingresos de \$12.000 o más.

Cabe destacar que la variable sobre consumo de bienes no resultó significativa para el análisis de diferenciación de la muestra según el análisis

<sup>4</sup> Dicho porcentaje surge utilizando la corrección propuesta por Greenacre (2008). Utilizando la corrección propuesta por Benzécri (1979), la varianza

explicada con los primeros dos factores aumenta a un 89%.

factorial. En este sentido, cabe reflexionar sobre la incapacidad de la misma para dar cuenta de la variabilidad de situaciones ya sea porque el consumo se haya extendido a distintos segmentos de la población como también porque la diferenciación en cuanto al mismo precise ser relevada con preguntas más específicas (por ejemplo, preguntando por productos menos extendidos y/o por marcas y modelos).

Por otro lado, los casos con valores faltantes tienden a ubicarse cercanos al centro de la distribución lo cual se condice con la caracterización previa que no mostraba diferencias significativas entre dichos valores.

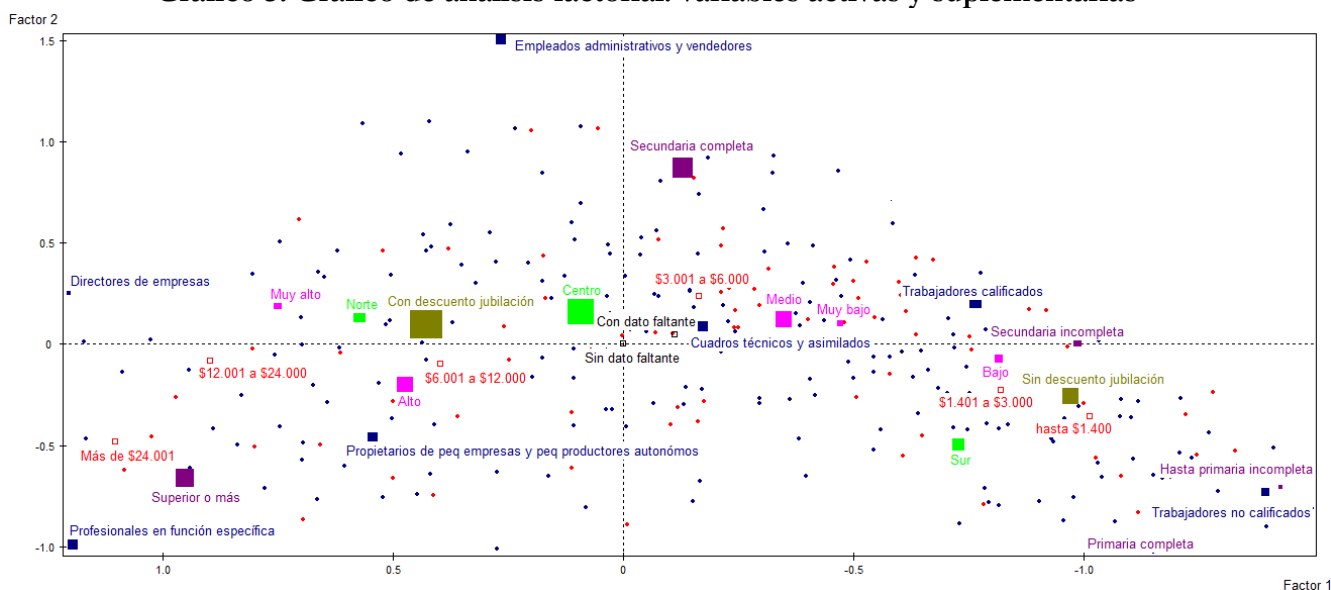
A partir de la visualización se observa que la variable clase social (con excepción de la categoría de cuadros técnicos y asalariados) tiende a corresponderse con la variable de rango de ingresos, es decir, que las clases más altas se ubican de forma cercana a los mayores ingresos mientras que las más bajas lo hacen con los menores ingresos. A su vez, la distribución de ambas variables sigue a la forma de la curva de Guttman (Baranger, 2009: 110) lo que da cuenta de determinado ordenamiento de las categorías. Mientras que la variable de nivel educativo utilizada frecuentemente como proxy de nivel socioeconómico se distribuye de forma más polarizada entre quienes tienen nivel superior o más y las restantes categorías, aunque también

presenta una forma de curva similar a las anteriores.

A continuación se realiza el análisis de clasificación de individuos en grupos mediante el procedimiento RECIP del programa SPAD que utiliza el método Ward y optimiza los resultados al hacer un ajuste por el método de centros móviles. De esta manera, se busca que los grupos resulten lo más heterogeneos entre sí y homogéneos al interior. Se obtienen cinco grupos que se caracterizan de la siguiente manera (ver Tablas 13, 14, 15, 16 y 17 del Anexo y Gráficos 4 y 5):

- 1) “Directivos”: directores de empresas que principalmente residen en zona norte.
  - 2) “Profesionales y pequeños propietarios”: profesionales en función específica y propietarios de pequeñas empresas, con descuento jubilatorio, con nivel educativo superior o mayor.
  - 3) “Trabajadores calificados”: Trabajadores administrativos, vendedores y técnicos, con hasta secundaria completa, con descuento jubilatorio y que residen principalmente en la zona norte.
  - 4) “Trabajadores no calificados”: Trabajadores sin calificación, con hasta primaria incompleta y que residen en zona sur.
- “Trabajadores informales”: Trabajadores sin descuentos jubilatorios, con hasta primaria completa y que residen en zona sur.

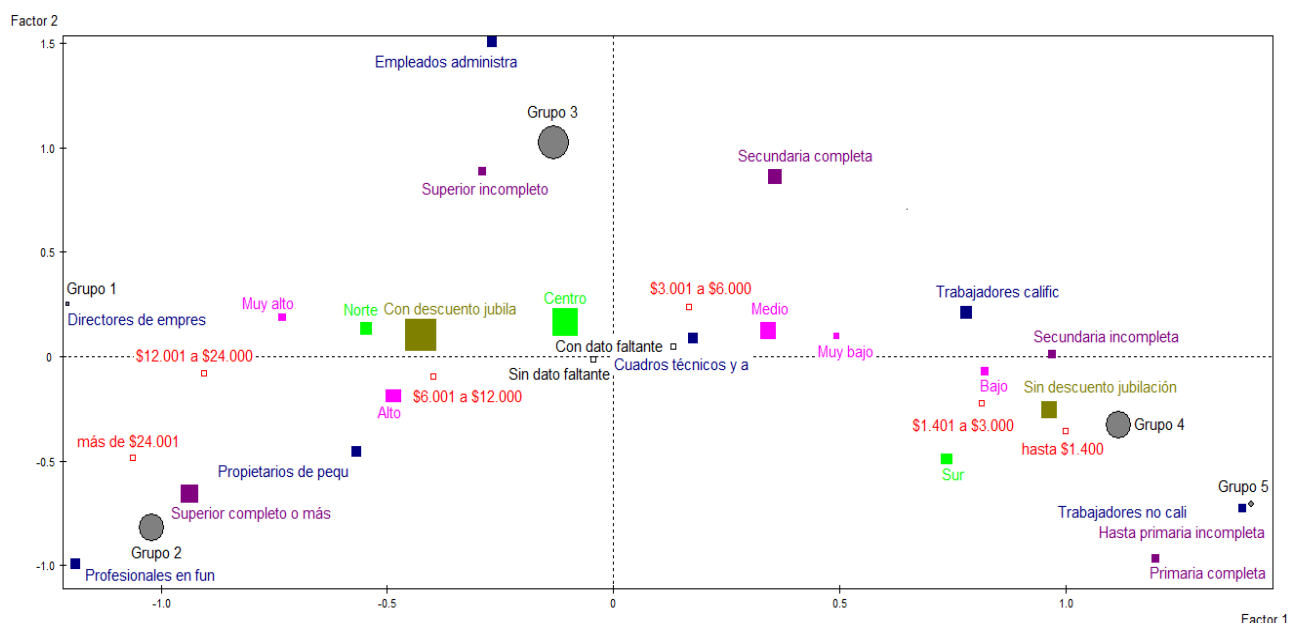
**Gráfico 5. Gráfico de análisis factorial: variables activas y suplementarias**



Fuente: elaboración propia en base a encuesta FONCYT 2012-2013



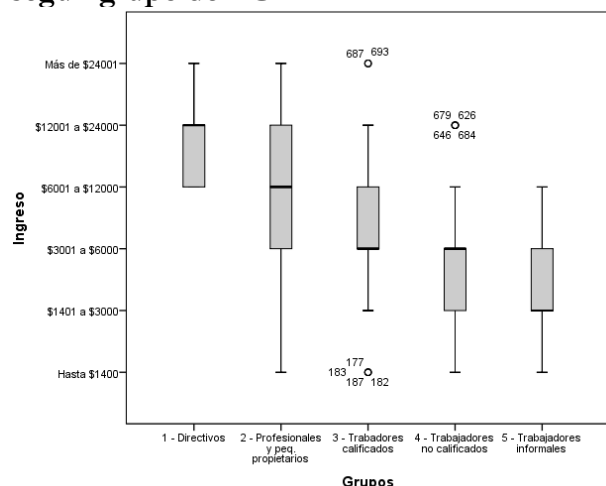
**Gráfico 5. Gráfico de análisis factorial y clasificación: variables activas, suplementarias y clusters**



Fuente: elaboración propia en base a encuesta FONCYT 2012-2013

A continuación, se observan las distribuciones de la variable ingreso para cada grupo conformado a partir del análisis anterior (Gráfico 6).

**Gráfico 6. Diagramas de caja de ingreso según grupo de ACM**



Fuente: elaboración propia en base a encuesta FONCYT 2012-2013

A partir de la conformación de dichos grupos se imputaron las medianas de ingresos siguiendo el siguiente cuadro (Tabla 7) donde se observa una distribución de ingresos similar al presentado por el esquema de clases de Torrado.

**Tabla 7. Medianas de ingreso (en intervalos) según grupos ACM**

Grupos	Mediana
1 - Directivos	\$12.001 a \$24.000
2 - Profesionales y pequeños propietarios	\$6.001 a \$12.000
3 - Trabajadores calificados	\$3.001 a \$6.000
4 - Trabajadores no calificados	\$1.401 a \$3.000
5 - Trabajadores informales	\$3.001 a \$6.000

Fuente: elaboración propia en base a Encuesta FONCYT 2012-2013

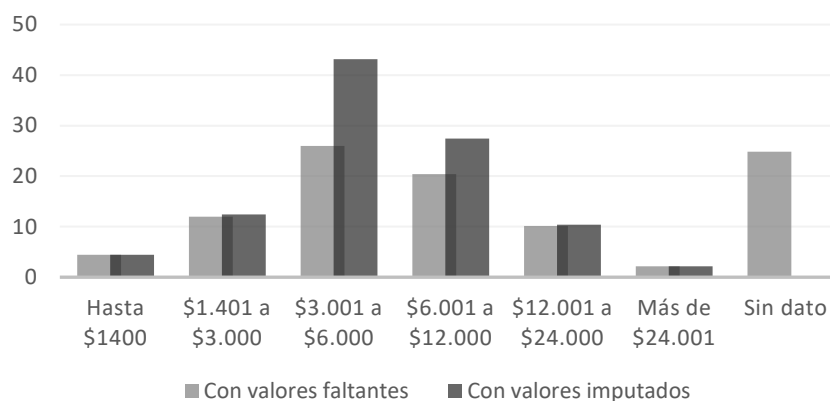
Por último, se realizó la imputación siguiendo la información anterior con los resultados del Gráfico 7.

Al compararse los resultados según cada método se observa que las dos categorías de los extremos de la variable de ingresos (hasta \$1.400 y más de \$24.001) y la anteúltima (\$12.001 a \$24.000) permanecen inalteradas mientras que las categorías intermedias aumentan su proporción de casos. Ambos tipos de imputación asignan alrededor de un 60% a la categoría de entre \$3.001 y \$6.000, siendo mayor el porcentaje en el caso del ACM/ACL, y luego, un 30% aproximadamente a la categoría de entre \$6.001 a

\$12.000, siendo mayor en el caso de la clase social. Además, se observa que mientras el método de la clase social otorga un 10% de casos con información faltante a la categoría de \$1.4001 a \$3.000, en el ACM ese porcentaje es sólo del 1,7%. En este sentido, se observa que el

ACM/ACL distribuye mayor cantidad de casos sin información en la categoría de \$3.001 a \$6.000 mientras que si se toma la clase social una parte de esos casos se reparten en las categorías aledañas de \$6.001 a \$12.000 y de \$1.4001 a \$3.000 (Tabla 8).

**Gráfico 7. Porcentaje de ingresos según muestra con valores faltantes y con valores imputados**



Fuente: elaboración propia en base a Encuesta FONCYT 2012-2013

**Tabla 8. Porcentaje de individuos según ingresos y método de imputación seleccionado**

Ingresos	Método de imputación			
	Clase social		ACM/ACL	
	Dato faltante	Total	Dato faltante	Total
Hasta \$1400	0,0	4,43	0,0	4,43
\$1.401 a \$3.000	9,8	14,43	1,7	12,43
\$3.001 a \$6.000	57,5	40,29	69,0	43,14
\$6.001 a \$12.000	31,6	28,29	28,2	27,43
\$12.001 a \$24.000	1,1	10,43	1,1	10,43
Más de \$24.001	0,0	2,14	0,0	2,14

## 6. Consideraciones finales

El presente trabajo tuvo como objeto mostrar y comparar dos formas de imputación de información faltante sobre ingresos, por un lado, mediante la utilización de la clase social de pertenencia y, por otro, con la aplicación de técnicas de análisis multivariado que permiten obtener agrupamientos de la población en base a distintos atributos considerados relevantes. A partir de la observación de las semejanzas y las diferencias entre ambas maneras de abordar esta cuestión se busca contribuir al problema frecuente de la ausencia de respuesta en la variable sobre ingresos en las encuestas.

En primer lugar, se tomó el esquema de estratos socio-ocupacionales elaborado por Susana Torrado y se obtuvo la distribución de los casos en los distintos estratos. Luego, se aplicó un procedimiento de imputación simple de las medianas de ingreso de la clase social a la que pertenecía cada uno de los casos con información faltante. Dicha elección se amparaba en el hecho de considerar a la clase social como un factor relevante a la hora de observar las condiciones de vida de los individuos y los hogares, principalmente a partir del nivel de ingresos obtenido.

En segundo lugar, se planteó un método de imputación a partir de la combinación del análisis de correspondencias múltiples y el análisis de clasificación. Lo que se buscó es, a partir de una construcción del espacio social compuesto por diversas dimensiones tales como el consumo, el lugar de residencia, el nivel de formalidad, el nivel educativo y la clase social, identificar en donde se posicionaban los individuos con información de ingresos monetarios faltante y qué características podían asumir, en función de su posicionamiento en dicho espacio.

En este sentido, a partir del análisis de clasificación se construyeron 5 grupos y se “transfirieron” las medianas del ingreso a los individuos de dichos conglomerados para los cuales no existía información. De esta manera, se pudo estratificar la población en base a una batería mayor de atributos vinculados al nivel socioeconómico de la población lo que podría considerarse una ventaja a la hora de tener que reconstruir la información sobre ingresos.

El análisis de correspondencias múltiples reafirma la centralidad de la clase social como variable de diferenciación entre la población, pero también concibe también como relevantes la zona de residencia y la cobertura jubilatoria.

En términos generales, las dos formas de imputación utilizadas clasificaron a los casos con datos faltantes de forma similar, aunque presentado algunas diferencias para los sujetos de ingresos medios y medios-bajos. En ambos casos, la asignación se realiza preponderantemente en las categorías intermedias de la variable de ingresos, sin embargo, se observa que el ACM/ACL presenta una relativa mayor concentración de los casos con información faltante que el procedimiento a partir de la clase social y que éste último asigna más cantidad de casos a una de las categorías bajas de ingresos.

A modo de síntesis, el análisis arroja como resultado que en ambas maneras de imputación la mayor parte de los datos faltantes se asignan en las categorías intermedias en términos de ingreso y además que la posición en la estructura socio-ocupacional constituye un elemento relevante para poder solucionar esa limitación de las

fuentes de datos. Aunque del análisis de correspondencias múltiples y de clasificación surge que también es pertinente incluir la zona de residencia y la cobertura jubilatoria como elementos fuertes de diferenciación social. Sin embargo, la imputación del ingreso a partir de los estratos socio-ocupacionales —elaborados por Torrado— tiende a subestimar los ingresos faltantes en relación con aquella que se realiza mediante los agrupamientos resultantes del análisis multivariado. En este sentido, y considerando que generalmente la no declaración se ubica entre aquellos de mayores ingresos (INDEC, 2015) y en posiciones de mayor jerarquía en términos de estratificación (Donza, 2011), se considera que la imputación a partir de técnicas multivariadas se revela con potencial para dar respuesta a una problemática tan acuciante como la falta de datos. Sin embargo, este trabajo es una primera aproximación que requiere seguir profundizándose para poder arribar a elaboraciones más concluyentes.

## 7. Bibliografía

- Acock, A. C.; Demo, D. H. (1994). *Family diversity and well-being*. New York.
- Baranger, D. (2009). *Construcción y análisis de datos. Introducción al uso de técnicas cuantitativas en la investigación social*. Recuperado a partir de <http://www.academia.edu/download/33093559/LibroDBconanda0a7.v2.pdf>
- Clemenceau, L.; Fernández Melián, M. C.; Rodríguez de la Fuente, J. (2016). Análisis de esquemas de clasificación social basados en la ocupación desde una perspectiva teórica-metodológica comparada. *Documentos de Jóvenes Investigadores IIGG*, 44.
- Donza, E. (2011), Incidencia de la no respuesta a las preguntas de ingresos en la Encuesta Permanente de Hogares. Consideraciones teóricas y efectos. Gran Buenos Aires 1990-2010, ponencia presentada en *IX Jornadas de Sociología de la Universidad de Buenos Aires*, 18/04/2015, disponible en [www.uca.edu.ar/uca/common/grupo68/files/jornada-edu-ppt.pdf](http://www.uca.edu.ar/uca/common/grupo68/files/jornada-edu-ppt.pdf)
- Echeverría Zabalza, J. (1999). *La movilidad social en España*. Istmo, Madrid.

- Erikson, R.; Goldthorpe, J. H. (1992). *The constant flux: A study of class mobility in industrial societies*. Oxford University Press, USA.
- Fachelli, S. (2012) Desigualdad y estratificación social en la Argentina, en Fachelli, Sandra; López, Néstor; López Roldán, Pedro y Sorrouille, Florencia. *Desigualdad y diversidad en América Latina: hacia un análisis tipológico comparado*. Buenos Aires: SITEAL, Instituto Internacional de Planeamiento de la Educación (UNESCO-OEI). Libros digitales, 2.
- Galván, M.; Medina, F. (2007). *Imputación de datos: teoría y práctica*. Santiago de Chile: Naciones Unidas, CEPAL, Div. de Estadística y Proyecciones Económicas. Recuperado a partir de <http://www.cepal.org/publicaciones/xml/9/29949/LCL2772e.pdf>
- Greenacre, M. (2007). *Correspondence analysis in practice*. CRC press.
- INDEC (2015). Ponderación de la muestra y tratamiento de valores faltantes en las variables de ingreso en la Encuesta Permanente de Hogares. *Metodología* No 15. Buenos Aires: INDEC.
- Le Roux, B.; Rouanet, H. (2010). *Multiple correspondence analysis*. Quantitative applications in the social sciences. Thousand Oaks, Calif: Sage Publications.
- López Roldán, P.; Fachelli, S. (2015). *Metodología de la Investigación Social Cuantitativa*. Bellaterra, Barcelona: UAB.
- López-Roldán, P. (2012). La construcción de tipologías para la medición de las desigualdades en Fachelli, Sandra; López, Néstor; López Roldán, Pedro y Sorrouille, Florencia. *Desigualdad y diversidad en América Latina: hacia un análisis tipológico comparado*. Buenos Aires: SITEAL, Instituto Internacional de Planeamiento de la Educación (UNESCO-OEI). Libros digitales, 2.
- Mazzeo, V.; Roggi, M. C. (2012). Los habitantes de hoteles familiares, pensiones, inquilinatos y casas tomadas de la Ciudad de Buenos Aires: ¿dónde están?, ¿de dónde vienen? ¿quiénes son? y ¿cómo viven? *Población de Buenos Aires*, 9(15), 7–28.
- Pérez López, C. (2004). Técnicas de análisis multivariante de datos. *Aplicaciones con SPSS, Madrid, Universidad Complutense de Madrid*.
- Pla, J. (2016). *Condiciones objetivas y esperanzas subjetivas. Movilidad social y marcos de (in)certidumbre. Un abordaje multidimensional de las trayectorias de clase. Argentina durante la primera década del siglo XXI*. Buenos Aires: Editorial Autores de Argentina.
- Pla, J., Rodríguez de la Fuente, J.; Fernández Melián, M. C. (2016). ¿Ascenso social o movilidad espuria?: Un análisis de las trayectorias de movilidad social. Argentina 2007-2008. *Temas y Debates*, (31), 99-122.
- Sacco, N. (2016). Las clases sociales en la Argentina según los censos de población de 1991 y 2001. *Revista Argentina de Estadística Aplicada*, 3, 1-17.
- Torrado, S. (1992). *Estructura social de la Argentina, 1945-1983*. Buenos Aires: Ediciones de la Flor.
- Torrado, S. (1998). *La medición empírica de las clases sociales. Familia y diferenciación social*. Eudeba.
- Vallejo, P. M. (2012). Introducción al Análisis de Varianza.

8. Anexo

**Tabla 9. Variables seleccionadas de esquemas: valores propios corregidos y porcentaje de inercia explicada\***

Factor	Valor propio	% Inercia	Valor propio corregido (*)	% Inercia (1)	% Inercia (2)	% Acumulado
1	0,449600	12,49%	0,097344	88,70%	73,64%	73,64%
2	0,266537	7,40%	0,006917	6,30%	5,23%	78,87%
3	0,234561	6,52%	0,001866	1,70%	1,41%	80,28%
4	0,231797	6,44%	0,001580	1,44%	1,20%	81,48%

Nota: \*1 según Benzécri y 2 según Greenacre.

Fuente: elaboración propia en base a encuesta FONCYT 2012-2013.

**Tabla 10. Contribuciones absolutas de las categorías de las variables originales a los ejes factoriales**

Variable	Peso relativo	Distancia al origen	Eje factorial 1
<b>Esquema de clase social Torrado</b>			
Directores de empresas	0,457	42,75000	1,48
Profesionales en función específica	3,171	5,30631	9,96
Propietarios de pequeñas empresas y pequeños productores autónomos	3,343	4,98291	2,40
Cuadros técnicos y asimilados	3,057	5,54206	0,21
Empleados administrativos y vendedores	3,200	5,25000	0,51
Trabajadores calificados	4,714	3,24242	6,39
Trabajadores no calificados	2,057	8,72222	8,85
<b>Consumo de bienes</b>			
Muy bajo	1,857	9,76923	1,01
Bajo	2,686	6,44681	4,02
Medio	6,143	2,25581	1,59
Alto	6,486	2,08370	3,40
Muy alto	2,829	6,07071	3,37
<b>Zonas de residencia</b>			
Norte	4,229	3,72973	2,82
Centro	11,029	0,81347	0,28
Sur	4,743	3,21687	5,73
<b>Descuento jubilatorio</b>			
Con descuento	13,914	0,43737	5,54
Sin descuento	6,086	2,28639	12,66
<b>Nivel educativo</b>			
Hasta primaria incompleta	0,714	27,00000	3,15
Primaria completa	2,086	8,58904	6,66
Secundaria incompleta	2,057	8,72222	4,31
Secundaria completa	5,257	2,80435	1,49
Superior incompleto	2,914	5,86275	0,54
Superior completo o más	6,971	1,86885	13,61

Fuente: elaboración propia en base a encuesta FONCYT 2012-2013.



**Tabla 11. Contribuciones relativas de las categorías de las variables originales a los ejes factoriales**

<b>Variables</b>	<b>Peso relativo</b>	<b>Distancia al origen</b>	<b>Eje factorial 1</b>
<b>Esquema de clase social Torrado</b>			
Directores de empresas	0,457	42,75000	0,03
Profesionales en función específica	3,171	5,30631	0,27
Propietarios de pequeñas empresas y pequeños productores autónomos	3,343	4,98291	0,06
Cuadros técnicos y asimilados	3,057	5,54206	0,01
Empleados administrativos y vendedores	3,200	5,25000	0,01
Trabajadores calificados	4,714	3,24242	0,19
Trabajadores no calificados	2,057	8,72222	0,22
<b>Consumo de bienes</b>			
Muy bajo	1,857	9,76923	0,02
Bajo	2,686	6,44681	0,10
Medio	6,143	2,25581	0,05
Alto	6,486	2,08370	0,11
Muy alto	2,829	6,07071	0,09
<b>Zonas de residencia</b>			
Norte	4,229	3,72973	0,08
Centro	11,029	0,81347	0,01
Sur	4,743	3,21687	0,17
<b>Descuento jubilatorio</b>			
Con descuento	13,914	0,43737	0,41
Sin descuento	6,086	2,28639	0,41
<b>Nivel educativo</b>			
Hasta primaria incompleta	0,714	27,00000	0,07
Primaria completa	2,086	8,58904	0,17
Secundaria incompleta	2,057	8,72222	0,11
Secundaria completa	5,257	2,80435	0,05
Superior incompleto	2,914	5,86275	0,01
Superior completo o más	6,971	1,86885	0,47

Fuente: elaboración propia en base a encuesta FONCYT 2012-2013.

**Tabla 12. Test-values de las categorías activas y suplementarias**

<b>VARIABLES Y CATEGORÍAS</b>	<b>CASOS</b>	<b>PESO ABSOLUTO</b>	<b>DISTANCIA AL ORIGEN</b>	<b>EJE 1</b>
<b>Esquema de clase social Torrado</b>				
Directores de empresas	16	16,00	42,75000	-4,88
Profesionales en función específica	111	111,00	5,30631	-13,64
Propietarios de pequeñas empresas y pequeños productores autónomos	117	117,00	4,98291	-6,73
Cuadros técnicos y asimilados	107	107,00	5,54206	1,98
Empleados administrativos y vendedores	112	112,00	5,25000	-3,10
Trabajadores calificados	165	165,00	3,24242	11,46
Trabajadores no calificados	72	72,00	8,72222	12,45
<b>Consumo de bienes</b>				
Muy bajo	65	65,00	9,76923	4,18
Bajo	94	94,00	6,44681	8,55
Medio	215	215,00	2,25581	6,01
Alto	227	227,00	2,08370	-8,89
Muy alto	99	99,00	6,07071	-7,85
<b>Zonas de residencia</b>				
Norte	148	148,00	3,72973	-7,49
Centro	386	386,00	0,81347	-3,14
Sur	166	166,00	3,21687	10,87
<b>Descuento jubilatorio</b>				
Con descuento	485	485,00	0,44330	-16,89
Sin descuento	212	212,00	2,30189	16,80
<b>Nivel educativo</b>				
Hasta primaria incompleta	25	25,00	27,00000	7,17
Primaria completa	73	73,00	8,58904	10,81
Secundaria incompleta	72	72,00	8,72222	8,69
Secundaria completa	184	184,00	2,80435	5,63
Superior incompleto	102	102,00	5,86275	-3,16
Superior completo o más	244	244,00	1,86885	-18,12
<b>Ingreso</b>				
hasta \$1.400	31	31,00	21,58060	5,69
\$1.401 a \$3.000	101	101,00	5,93069	8,85
\$3.001 a \$6.000	282	282,00	1,48227	3,66
\$6.001 a \$12.000	198	198,00	2,53535	-6,62
\$12.001 a \$24.000	73	73,00	8,58904	-8,16
más de \$24.001	15	15,00	45,66670	-4,16
<b>Datos faltantes</b>				
Con dato faltante	174	174,00	3,02299	2,02
Sin dato faltante	526	526,00	0,33080	-2,02

Fuente: elaboración propia en base a encuesta FONCYT 2012-2013.

**Tabla 13. Caracterización del grupo 1**

Variable	Categoría	Test-value	Probabilidad	Peso
Clase social	Directores de empres	11,88	0,000	16
Ingreso	\$12.001 a \$24.000	3,84	0,000	73
Zona de residencia	Norte	2,36	0,009	148
Ingreso	\$3.001 a \$6.000	-3,50	0,000	282
Valores perdidos	Con datos perdidos	-99,99	0,000	0

Fuente: elaboración propia en base a encuesta FONCYT 2012-2013.

**Tabla 14 Caracterización del grupo 2**

Variable	Categoría	Test-value	Probabilidad	Peso
Nivel educativo	Superior completo o	22,22	0,000	244
Clase social	Profesionales en fun	17,11	0,000	111
Descuento jubilatorio	Con descuento jubila	8,53	0,000	485
Clase social	Propietarios de pequ	6,55	0,000	117
Consumo de bienes	Alto	5,24	0,000	227
Ingreso	\$6.001 a \$12.000	5,17	0,000	198
Ingreso	\$12.001 a \$24.000	3,86	0,000	73
Consumo de bienes	Muy alto	3,26	0,001	99
Ingreso	más de \$24.001	2,74	0,003	15
Ingreso	hasta \$1.400	-2,47	0,007	31
Clase social	Directores de empres	-2,72	0,003	16
Nivel educativo	Hasta primaria incom	-3,67	0,000	25
Consumo de bienes	Medio	-3,73	0,000	215
Ingreso	\$3.001 a \$6.000	-4,15	0,000	282
Consumo de bienes	Bajo	-4,27	0,000	94
Zona de residencia	Sur	-4,63	0,000	166
Ingreso	\$1.401 a \$3.000	-4,69	0,000	101
Nivel educativo	Superior incompleto	-5,32	0,000	102
Nivel educativo	Secundaria incomplet	-6,97	0,000	72
Clase social	Trabajadores no cali	-6,97	0,000	72
Nivel educativo	Primaria completa	-7,02	0,000	73
Descuento jubilatorio	Sin descuento jubila	-8,39	0,000	212
Clase social	Trabajadores calific	-8,56	0,000	165
Clase social	Empleados administra	-9,03	0,000	112
Nivel educativo	Secundaria completa	-11,13	0,000	184
Valores perdidos	Con datos perdidos	-99,99	0,000	0

Fuente: elaboración propia en base a encuesta FONCYT 2012-2013.

**Tabla 15 Caracterización del grupo 3**

Variable	Categoría	Test-value	Probabilidad	Peso
Clase social	Empleados administra	15,09	0,000	112
Nivel educativo	Superior incompleto	10,79	0,000	102
Nivel educativo	Secundaria completa	9,54	0,000	184
Descuento jubilatorio	Con descuento jubila	5,84	0,000	485
Ingreso	\$3.001 a \$6.000	3,44	0,000	282
Zona de residencia	Norte	2,97	0,001	148
Clase social	Cuadros técnicos y a	2,58	0,005	107
Ingreso	hasta \$1.400	-2,61	0,005	31
Ingreso	\$1.401 a \$3.000	-2,77	0,003	101
Clase social	Directores de empres	-3,12	0,001	16
Zona de residencia	Sur	-3,78	0,000	166
Nivel educativo	Hasta primaria incom	-4,16	0,000	25
Nivel educativo	Secundaria incomplet	-4,72	0,000	72
Nivel educativo	Primaria completa	-5,45	0,000	73
Clase social	Trabajadores no cali	-5,72	0,000	72
Descuento jubilatorio	Sin descuento jubila	-5,86	0,000	212
Nivel educativo	Superior completo o	-10,02	0,000	244
Clase social	Profesionales en fun	-10,04	0,000	111
Valores perdidos	Con datos perdidos	-99,99	0,000	0

Fuente: elaboración propia en base a encuesta FONCYT 2012-2013.

**Tabla 16. Caracterización del grupo 4**

Variable	Categoría	Test-value	Probabilidad	Peso
Nivel educativo	Hasta primaria incom	99,99	0,000	25
Ingreso	\$1.401 a \$3.000	3,92	0,000	101
Clase social	Trabajadores no cali	2,85	0,002	72
Consumo de bienes	Bajo	2,73	0,003	94
Descuento jubilatorio	Sin descuento jubila	2,53	0,006	212
Zona de residencia	Sur	2,51	0,006	166
Clase social	Propietarios de pequ	-2,35	0,009	117
Descuento jubilatorio	Con descuento jubila	-2,48	0,007	485
Consumo de bienes	Alto	-2,65	0,004	227
Nivel educativo	Secundaria completa	-3,34	0,000	184
Nivel educativo	Superior completo o	-4,14	0,000	244
Valores perdidos	Con datos perdidos	-99,99	0,000	0

Fuente: elaboración propia en base a encuesta FONCYT 2012-2013.

**Tabla 17. Caracterización del grupo 5**

Variable	Categoría	Test-value	Probabilidad	Peso
Descuento jubilatorio	Sin descuento jubila	13,21	0,000	212
Nivel educativo	Primaria completa	11,86	0,000	73
Nivel educativo	Secundaria incomplet	10,85	0,000	72
Clase social	Trabajadores no cali	9,74	0,000	72
Clase social	Trabajadores calific	9,68	0,000	165
Zona de residencia	Sur	7,55	0,000	166
Ingreso	\$1.401 a \$3.000	5,65	0,000	101
Ingreso	hasta \$1.400	5,12	0,000	31
Consumo de bienes	Bajo	3,77	0,000	94
Consumo de bienes	Medio	3,65	0,000	215
Consumo de bienes	Muy bajo	3,17	0,001	65
Ingreso	más de \$24.001	-2,56	0,005	15
Clase social	Directores de empres	-2,68	0,004	16
Nivel educativo	Hasta primaria incom	-3,63	0,000	25
Clase social	Propietarios de pequ	-4,69	0,000	117
Ingreso	\$6.001 a \$12.000	-5,07	0,000	198
Consumo de bienes	Alto	-5,12	0,000	227
Ingreso	\$12.001 a \$24.000	-5,16	0,000	73
Consumo de bienes	Muy alto	-5,36	0,000	99
Zona de residencia	Norte	-6,34	0,000	148
Nivel educativo	Superior incompleto	-6,48	0,000	102
Clase social	Profesionales en fun	-8,02	0,000	111
Clase social	Empleados administra	-8,07	0,000	112
Nivel educativo	Superior completo o	-11,73	0,000	244
Descuento jubilatorio	Con descuento jubila	-13,38	0,000	485
Valores perdidos	Con datos perdidos	-99,99	0,000	0

Fuente: elaboración propia en base a encuesta FONCYT 2012-2013.